Hepatoprotective effects of systemic ER activation

ChIPseq/Epigenome genome - Quantification of reads in diffbound promoters and enhancers

Christian Sommerauer & Carlos Gallardo

25 August, 2022

This script has two parts:

in Part 1, we use thr generated bed file with 182 diffbound promoters (one TSS per gene; the file after diffbind analysis still has multiple TSS per gene, 184 total). This bedfile is then converted into a SAF file using awk in command line. This SAF file, together with BAM files is used for featurecounts to generate a count matrix with all reads in these peak regions (this is analyzed on UPPMAX). The same is done for the 1,816 enhancer peaks.

in Part 2, we import this count matrix into R again, normalize it, calculate significances and generate the plots.

Part 1

```
#The enhancer and promoter (with unique TSS, generated in the chunk above) bed files were used to creat

awk 'OFS="\t" {print $1"."$2"."$3, $1, $2, $3, "."}' in.bed > out.saf

#The saf file was then used to annotate the reads, all bam files (Remove duplicates, blocklist, sorted,
#subread version 2.0.0 was used.

Following featureCounts chunk was used:
featureCounts \
        -g gene_id \
        -s 2 \
        -T 2 \
        -M \
        -F SAF \
        -O \
        -C \
        -a ${SAF_PATH}/example.saf  \
        -o ${OUTPUT_PATH}/example.readCount \
        ${BAM_PATH}/*H3K27ac*MkDup.bam \
        &> ${OUTPUT_PATH}/example.readCount.log

the code was then uploaded to uppmax, features counted, then downloaded, finally the header cleaned up
```

Part 2

Import the reads in peaks which was created using feature counts. Then, normalize the reads to do the analysis. Perform an Anova to compare significances between the conditions.

```
library(tidyverse)
counts_prom <- read.delim("results/Epigenome_analysis/diffbind_promoters_182_H3K27ac.clean.readCount", h
names(counts_prom) <- c("CDm2","CDm9","HFDm3","HFDm4", "DPN2","DPN3","E2_8","E2_9")
colsums_prom <- colSums(counts_prom[,])

#normalise per depth
counts_prom_norm <- sweep(counts_prom, 2, colsums_prom, FUN = "/")
counts_prom_norm2 <- counts_prom_norm *10^6
colSums(counts_prom_norm2[,])
```

First, for promoters.

```
##  CDm2  CDm9 HFDm3 HFDm4  DPN2  DPN3  E2_8  E2_9
## 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06
```

```
#Combine the replicates
counts_prom_norm3 <- counts_prom_norm2 %>%
  mutate(avg_CD = rowMeans(counts_prom_norm2[,1:2])) %>%
  mutate(avg_HFD = rowMeans(counts_prom_norm2[,3:4])) %>%
  mutate(avg_DPN = rowMeans(counts_prom_norm2[,5:6])) %>%
  mutate(avg_E2 = rowMeans(counts_prom_norm2[,7:8])) %>%
  tibble::rownames_to_column("loc_ID")

prom_enh_diffbound <- readRDS("results/Epigenome_analysis/annotated_diffbind_and_genomewide_promoters_en

Prom_Diffbind_HFDup <- prom_enh_diffbound$promoters_HFDup %>% mutate("loc_ID" = paste0(seqnames, ".",sta
Prom_Diffbind_HFDdown <- prom_enh_diffbound$promoters_HFDdown %>% mutate("loc_ID" = paste0(seqnames, "."

counts_prom_HFDup <- counts_prom_norm3 %>% filter(loc_ID%in%Prom_Diffbind_HFDup$loc_ID) %>% dplyr::selec


counts_prom_HFDup2 <- counts_prom_HFDup %>%
  pivot_longer(cols=2:5) %>% mutate("updown"= "HFDup") %>% group_by(name)

# STATISTICS - One-sided anova.
counts_prom_HFDup.stat <- aov(value ~ name, data = counts_prom_HFDup2)
summary(counts_prom_HFDup.stat)
```

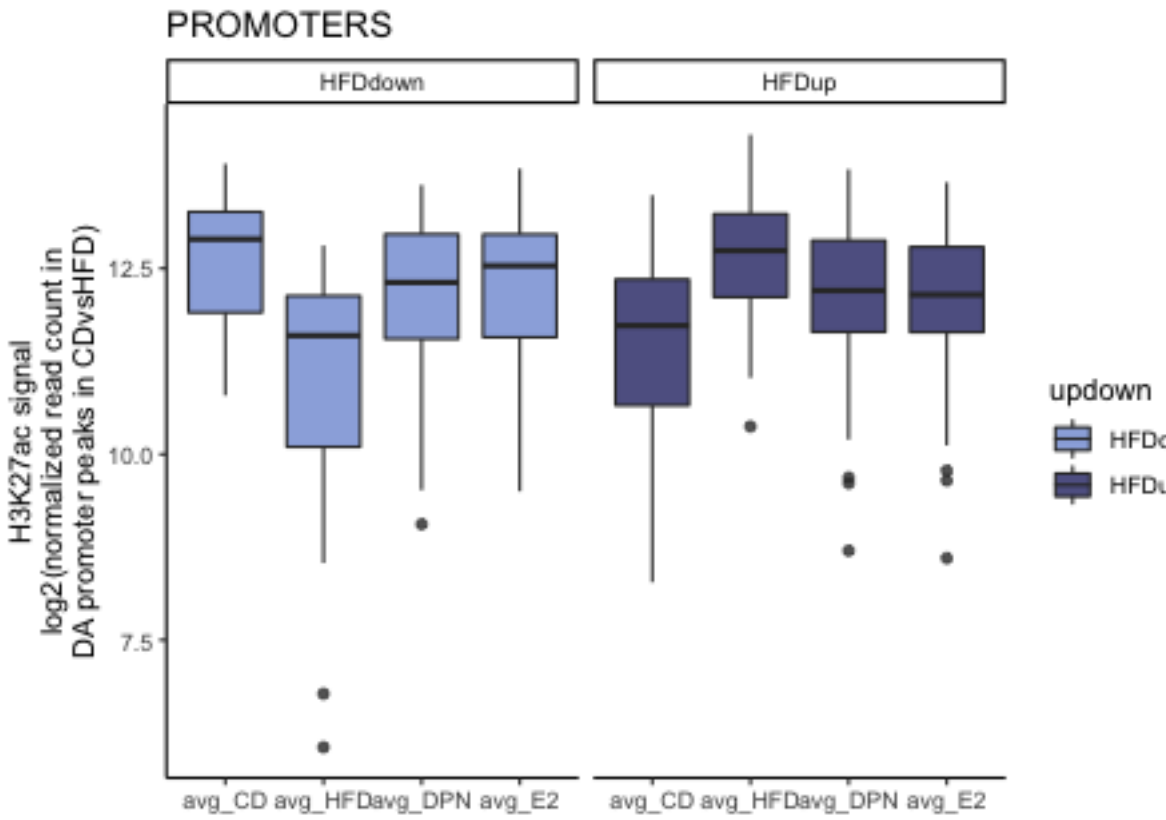
```
##              Df    Sum Sq   Mean Sq F value   Pr(>F)
## name          3 6.129e+08 204312689   21.19 8.74e-13 ***
## Residuals   412 3.972e+09   9641098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
counts_prom_HFDup.stat1 <- TukeyHSD(counts_prom_HFDup.stat)
counts_prom_HFDup.stat2 <- counts_prom_HFDup.stat1$name

counts_prom_HFDdown <- counts_prom_norm3 %>% filter(loc_ID%in%Prom_Diffbind_HFDdown$loc_ID) %>% dplyr::s
counts_prom_HFDdown2 <- counts_prom_HFDdown %>%
  pivot_longer(cols=2:5) %>% mutate("updown"="HFDdown") %>% group_by(name)

# STATISTICS - One-sided anova.
counts_prom_HFDdown.stat <- aov(value ~ name, data = counts_prom_HFDdown2)
summary(counts_prom_HFDdown.stat)
```

```
##              Df    Sum Sq   Mean Sq F value   Pr(>F)
## name          3 8.173e+08 272416919   27.98 5.02e-16 ***
## Residuals   308 2.999e+09   9737480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
counts_prom_HFDdown.stat1 <- TukeyHSD(counts_prom_HFDdown.stat)
counts_prom_HFDdown.stat2 <- counts_prom_HFDdown.stat1$name

counts_prom_HFD_plot <- rbind(counts_prom_HFDup2, counts_prom_HFDdown2)
counts_prom_HFD_plot$value <- counts_prom_HFD_plot$value+1
```

```
order <- c("avg_CD", "avg_HFD", "avg_DPN", "avg_E2")
ggplot(counts_prom_HFD_plot, aes(x=factor(name, levels=order), y=log2(value), fill=updown)) +
  geom_boxplot(alpha=0.8) +
  theme_classic() +
  facet_wrap(vars(updown)) +
  xlab("") +
  ggtitle("PROMOTERS") +
  ylab("H3K27ac signal\n log2(normalized read count in \nDA promoter peaks in CDvsHFD)") +
  scale_fill_manual(values = c("#7f9ad7", "#2c2f72"))
```



Plot for promoters

```
library(dplyr)
library(tidyr)
counts_enha <- read.delim("results/Epigenome_analysis/diffbind_enhancers_1816_H3K27ac.clean.readCount",
names(counts_enha) <- c("CDm2","CDm9","HFDm3","HFDm4", "DPN2","DPN3","E2_8","E2_9")
colsums_enha <- colSums(counts_enha[,])

counts_enha_norm <- sweep(counts_enha, 2, colsums_enha, FUN = "/")
counts_enha_norm2 <- counts_enha_norm *10^6
colSums(counts_enha_norm2[,])
```

Then, for enhancers

```
##  CDm2  CDm9 HFDm3 HFDm4  DPN2  DPN3  E2_8  E2_9
## 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06 1e+06
```

```
#Combine the replicates
counts_enha_norm3 <- counts_enha_norm2 %>%
  mutate(avg_CD = rowMeans(counts_enha_norm2[,1:2])) %>%
  mutate(avg_HFD = rowMeans(counts_enha_norm2[,3:4])) %>%
  mutate(avg_DPN = rowMeans(counts_enha_norm2[,5:6])) %>%
  mutate(avg_E2 = rowMeans(counts_enha_norm2[,7:8])) %>%
  tibble::rownames_to_column("loc_ID")
```

```r
#Import the diffbound enhancers (up, down separately)
Enha_Diffbind_HFDup <- prom_enh_diffbound$enhancers_HFDup %>% mutate("loc_ID" = paste0(seqnames, ".",sta
Enha_Diffbind_HFDdown <- prom_enh_diffbound$enhancers_HFDdown %>% mutate("loc_ID" = paste0(seqnames, "."

counts_enha_HFDup <- counts_enha_norm3 %>% filter(loc_ID%in%Enha_Diffbind_HFDup$loc_ID) %>% dplyr::sele
counts_enha_HFDup2 <- counts_enha_HFDup %>%
  pivot_longer(cols=2:5) %>% mutate("updown"= "HFDup")

# STATISTICS - One-sided anova.
res.aov.enha.HFDup.stat <- aov(value ~ name, data = counts_enha_HFDup2)
summary(res.aov.enha.HFDup.stat)
```

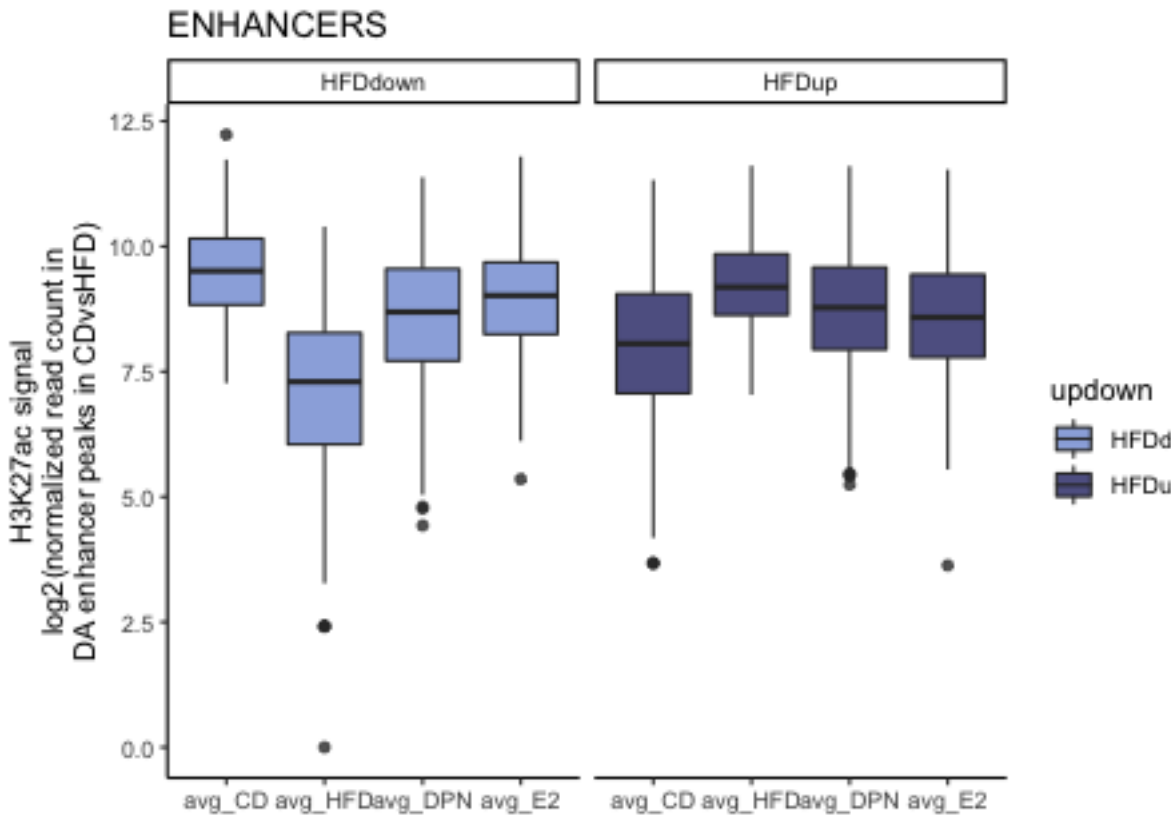
```
##               Df    Sum Sq  Mean Sq F value Pr(>F)
## name           3  65854235 21951412     137 <2e-16 ***
## Residuals   4840 775748560   160279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
res.aov.enha.HFDup.stat1 <- TukeyHSD(res.aov.enha.HFDup.stat)
res.aov.enha.HFDup.stat2 <- res.aov.enha.HFDup.stat1$name

counts_enha_HFDdown <- counts_enha_norm3 %>% filter(loc_ID%in%Enha_Diffbind_HFDdown$loc_ID) %>% dplyr::s
counts_enha_HFDdown2 <- counts_enha_HFDdown %>%
  pivot_longer(cols=2:5) %>% mutate("updown"="HFDdown")

# STATISTICS - One-sided anova.
res.aov.enha.HFDdown.stat <- aov(value ~ name, data = counts_enha_HFDdown2)
summary(res.aov.enha.HFDdown.stat)
```

```
##               Df    Sum Sq  Mean Sq F value Pr(>F)
## name           3 131817320 43939107   214.2 <2e-16 ***
## Residuals   2416 495700772   205174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
res.aov.enha.HFDdown.stat1 <- TukeyHSD(res.aov.enha.HFDdown.stat)
res.aov.enha.HFDdown.stat2 <- res.aov.enha.HFDdown.stat1$name

counts_enha_HFD_plot <- rbind(counts_enha_HFDup2, counts_enha_HFDdown2)
counts_enha_HFD_plot$value <- counts_enha_HFD_plot$value+1
```

```r
ggplot(counts_enha_HFD_plot, aes(x=factor(name, levels=order), y=log2(value), fill=updown)) +
  geom_boxplot(alpha=0.8) +
  theme_classic() +
  facet_wrap(vars(updown)) +
  xlab("") +
  ggtitle("ENHANCERS") +
  ylab("H3K27ac signal\n log2(normalized read count in \nDA enhancer peaks in CDvsHFD)") +
  scale_fill_manual(values = c("#7f9ad7", "#2c2f72"))
```



Plot for enhancers

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] forcats_0.5.1   stringr_1.4.0   dplyr_1.0.6     purrr_0.3.4
## [5] readr_1.4.0     tidyr_1.1.3     tibble_3.1.2    ggplot2_3.3.3
## [9] tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.1  xfun_0.31         haven_2.4.1       colorspace_2.0-1
##  [5] vctrs_0.3.8       generics_0.1.0    htmltools_0.5.1.1 yaml_2.2.1
##  [9] utf8_1.2.1        rlang_0.4.11      pillar_1.6.1      glue_1.6.2
## [13] withr_2.4.2       DBI_1.1.1         dbplyr_2.1.1      modelr_0.1.8
```

```
## [17] readxl_1.3.1      lifecycle_1.0.0   munsell_0.5.0     gtable_0.3.0
## [21] cellranger_1.1.0  rvest_1.0.0       evaluate_0.14     labeling_0.4.2
## [25] knitr_1.33        fansi_0.5.0       highr_0.9         broom_0.7.6
## [29] Rcpp_1.0.6        scales_1.1.1      backports_1.2.1   jsonlite_1.7.2
## [33] farver_2.1.0      fs_1.5.0          hms_1.1.0         digest_0.6.27
## [37] stringi_1.6.2     grid_4.0.3        cli_3.2.0         tools_4.0.3
## [41] magrittr_2.0.1    crayon_1.4.1      pkgconfig_2.0.3   ellipsis_0.3.2
## [45] xml2_1.3.2        reprex_2.0.0      lubridate_1.7.10  assertthat_0.2.1
## [49] rmarkdown_2.14    httr_1.4.2        rstudioapi_0.13   R6_2.5.0
## [53] compiler_4.0.3
```